Phylogenetic comparative methods: harnessing the power of species diversity to investigate welfare issues in captive wild animals

Short title: Phylogenetic comparative methods for zoos

Emma Mellor[1], Heather McDonald Kinkaid[2], Georgia Mason[2*]

[1]Bristol Veterinary School, University of Bristol, Langford, UK

[2]Department of Animal Biosciences, University of Guelph, Ontario, Canada

[*]Correspondence to: gmason@uoguelph.ca

Abstract

This paper reviews a way of investigating health and welfare problems in captive wild animals (*e.g.* those in zoos, aviaries, aquaria or aquaculture systems) that has great potential, but to date has been little used: systematically comparing species with few or no health and welfare issues to those more prone to problems. Doing so empirically pinpoints species-typical welfare risk and protective factors (such as aspects of their natural behavioral biology): information which can then be used to help prevent or remedy problems by suggesting new ways to improve housing and husbandry, and by identifying species intrinsically best suited to captivity. We provide a detailed, step-by-step 'how to' guide for researchers interested in using these techniques, including guidance on how to statistically

control for the inherent similarities shared by related species: an important concern because simple, cross-species comparisons that do not do this may well fail to meet statistical assumptions of non-independence. The few relevant studies that have investigated captive wild animals' welfare problems using this method are described. Overall, such approaches reap value from the great number and diversity of species held in captivity (*e.g*. the many thousands of species held in zoos); can yield new insights from existing data and published results; render previously intractable welfare questions (such as "do birds need to fly?" or "do Carnivora need to hunt?") amenable to study; and generate evidence-based principles for integrating animal welfare into collection planning.

Key words: species differences; welfare; independent contrasts; PGLS

Introduction

Ensuring good animal health and welfare is part of the ethos of modern zoos (AZA, 2018; EAZA, 2013; Hill & Broom, 2009; WAZA, 2018). Indeed, the World Association of Zoos and Aquariums (WAZA) states: "*zoos and aquariums have a responsibility to achieve high standards of animal welfare*" (Mellor, Hunt, & Gusset, 2015). This reflects moral and legal obligations that animals kept by humans should be well cared for (*e.g.* GOV.UK, 2013; Hill & Broom, 2009; RSPCA, 2018). But good welfare yields practical benefits too. For example, good animal welfare improves the public's perception of captive facilities (*e.g*. Miller, 2012), and also helps zoos meet their aims of achieving self-sustaining populations (Hosey, Melfi, & Pankhurst, 2013) by ensuring that as many individuals as required successfully mate and produce viable progeny. This is because poor welfare can compromise libido, fertility, parental care, and survivorship (*e.g.* Bronson *et al*., 2007; Díez-León *et al*., 2013; Mason, Leipoldt, & de Jonge, 1995; Peng *et al*., 2007).

Zoos keep a vast, diverse taxonomic array of animals. For terrestrial vertebrates alone, Species360 zoos hold nearly 4,000 species (Conde *et al*., 2013) and, therefore, species-specific, specialized research is often required to optimize husbandry. Zoo health and welfare researchers have three main research methods at their disposal, two of which are already commonplace. One is experimental. Here, the effects of experimentally providing a treatment are recorded, with subjects often acting as their own controls. Examples include studies of the effects of UV provision on broad-snouted caiman (*Caiman latirostris,* Daudin, 1802) (Siroski, Poletta, Fernandez, Ortega, & Merchant, 2012); of carotenoid supplementation on southern corroboree tadpoles (*Pseudophryne corroboree*, Moore, 1953) (Byrne & Silla, 2017); of dietary manipulations on lemurs (Britt, Cowlard, Baker, & Plowman, 2015); and myriad environmental enrichment studies (*e.g.* Schneider, Nogge, & Kolter, 2014; Wallace, Kingston-Jones, Ford, & Semple, 2013). Similar 'pseudo-experimental' research instead opportunistically studies the effects of non-experimental manipulations, such as the impacts of

visitor-generated noise (*e.g.* Quadros, Goulart, Passos, Vecci, & Young, 2014) and inter-zoo transfers (*e.g.* Schmid, Heistermann, Ganslosser, & Hodges, 2001; Snyder *et al*., 2012).

The second common research approach is epidemiological. Here, unplanned, pre-existing variation in various aspects of husbandry or health care is used in a between-subject approach (*e.g.* comparing animals across different enclosures or zoos). Examples include: Blay and Côté (2001)'s survey of enclosure-related effects on breeding and mortality in Humboldt penguins (*Spheniscus humboldti*, Meyen, 1834*);* research into effects of birth origin on the survivorship of zoo elephants (*Loxodonta africana*, Blumenbach, 1797; *Elephas maximus*, Linnaeus, 1758: Clubb *et al*., 2009); and recent multi-zoo studies of housing- and husbandry-related risk factors for stereotypic behavior in polar bears (*Ursus maritimus*, Phipps, 1774) (Shepherdson, Lewis, Carlstead, Bauman & Perrin, 2013) and elephants (Greco *et al*., 2016).

The third research approach is the focus of our paper: exploring the correlates of variation across different species (where 'species', rather than individual, enclosure, or zoo, is the unit of replication). Like epidemiological approaches, this methodology exploits pre-existing variation in health and welfare problems: here, variation between different species. It then identifies what makes some species prone to welfare problems in captivity, but others – sometimes even closely related species – instead resilient and apparently protected from such issues. Conservation biologists working on *in situ* populations have long used this approach to reveal why species differ in extinction risk, vulnerability to human exploitation, invasiveness, and other conservation-related attributes (*e.g.* Cardillo *et al*., 2005; Fisher & Owens, 2004), and it has huge potential for advancing the understanding of zoo animal welfare (Clubb & Mason, 2004; Mason, 2010). However, this approach has been relatively little used to date, perhaps because it requires formal "phylogenetic comparative methods" (PCMs): statistical methods permitting correct statistical comparison across species (*e.g.* Cornwell & Nakagawa, 2017).

The purpose of this review is therefore to provide an introduction to PCMs tailored for researchers interested in studying welfare problems in zoos, aquaria, and similar systems. First, we

first explain how, if used naïvely, research into between-species variation can fail to meet statistical assumptions of non-independence; why this matters; and how this problem can be solved statistically. We then describe the few studies to investigate captive wild animals' welfare problems using PCMs. Next we provide a step-by-step 'how to' guide to using these techniques, including how to statistically control for e non-independence of related species. To end, we discuss how this approach for understanding and improving animal welfare can complement other methodologies, and how it may even have unique value, making previously intractable questions tractable and providing principles to assist collection planning.

Why use "phylogenetic" comparative methods when comparing species?

The principle behind comparing species to test welfare-related hypotheses is quite simple. If, for example, one wanted to test the hypothesis that being able to fly is important for avian welfare, one would collect data on welfare indicators (*e.g.* stereotypic behavior, egg hatchability, chick mortality, or the prevalence of opportunistic infections, *cf.* Appleby, Olsson, & Galindo, 2018; Hill & Broom, 2009; Mason & Veasey, 2010) from a range of species that differ in reliance on flight in the wild. If being able to fly is important for captive bird welfare, this makes the testable prediction that naturally flightless species should have the best captive welfare (because they have no flight behavior to be constrained); while species that fly a lot, for example relying on flight to feed or migrate, should have the poorest welfare, because naturally strong flying motivations are frustrated. This thus predicts a positive relationship between metrics of species-typical reliance on flight in nature and species-typical captive welfare problems (see Fig. 1A).

However, the simple regression depicted in Fig. 1A is inappropriate. Recognized for decades (*e.g.* Clutton-Brock & Harvey, 1977), comparing species as though each is an independent datapoint

(a key assumption of most standard statistical tests) is problematic. This is because species are part of hierarchical structures (or 'phylogenies'), and so typically cannot be considered independent from each other (Felsenstein, 1985; Grafen, 1989; Harvey & Rambaut, 1998; Martins & Garland, 1991; Purvis & Rambaut, 1995). Such shared ancestry often results in non-independence, or pseudoreplication, because closely related species are likely to resemble each other, sharing similar evolved and non-evolved attributes (Harvey & Pagel, 1991). This similarity based on relatedness is termed 'phylogenetic signal' (Blomberg, Garland, & Ives, 2003; Grafen, 1989; Pagel, 1999). If this is ignored and standard statistical tests used, species are incorrectly assumed to be statistically independent (Diaz-Uriarte & Garland, 1996; though see: Revell, 2010), and phylogeny may confound the analysis.

FIGURE 1 HERE PLEASE

Why this matters is illustrated by the (fictitious) data in Fig. 1B. This reveals that the data shown in Fig. 1A come from two separate groups of closely related species which cluster together, effectively reducing our eight (pseudoreplicative) datapoints to just two groupings of similar birds. The flamingo species are all intrinsically similar to one another in their low reliance on flight, but also in being large-bodied, aquatic filter-feeders, dramatic-looking to human visitors, *etc*. Likewise, the swallows are all intrinsically similar to each other in their heavy reliance on flight, but also in being small-bodied, insectivorous, duller in appearance to humans, *etc*. Ignoring phylogeny thus makes it impossible to validly assess whether there is a correlation between daily time spent flying and signs of poor welfare, because *any* of the attributes that flamingos share with one another (and do not share with swallows) might equally explain the apparent relationship (*cf.* Cuthill, 2005). Our fictitious example thus reveals the regression in Fig. 1A to be a Type I error: there is no convincing evidence that constraints on flying predict poor welfare, because *within* each group, the relatively greater fliers do not have the poorest welfare. Thus after parsing out phylogenetic relatedness, different patterns can emerge from species data; and when they do, *these* are the ones that test hypotheses validly.

But how to parse out such relatedness statistically? Felsenstein (1985)'s seminal paper was the first to show how to statistically solve this problem by incorporating phylogenetic relationships between species into analyses. This paper presented a method called 'phylogenetic independent contrasts' (see "Data analysis and interpretation", below, for details), and thus 'phylogenetic comparative methods' (PCMs) were born. PCMs have since undergone rapid development, with various options now available, such 'phylogenetic generalized least squares regressions' (Grafen, 1989) (see "Data analysis and interpretation", below, for details). Are PCMs *essential* when comparing species to test hypotheses? The simple answer is yes, to avoid pseudoreplication. PCMs have thus robustly withstood criticism from some (*e.g.* Björklund, 1997; Westoby, Leishman, & Lord, 1995). Furthermore, not using PCMs to analyze species data can alter results. Simulation studies repeatedly demonstrate that PCMs out-perform standard statistical tests (*e.g.* Revell, 2010), reducing both Type I (*e.g.* Diaz-Uriarte & Garland, 1996) and Type II error rates, so increasing statistical power (Arnold, Matthews, & Nunn, 2010). PCMs are therefore widely accepted as the correct way to analyze species data when testing hypotheses (with Felsenstein's seminal paper since accruing over 5,500 citations, Web of Science [WoS] accessed 03/20/18), and are mainstream research tools for evolutionary biologists and behavioral ecologists (reviewed by Cornwell & Nakagawa, 2017; Freckleton, 2009).

PCMs and welfare issues in captive wild animals: an overview of past research

PCMs have started to be used to investigate captive wild animal welfare issues, testing hypotheses about risk factors by correlating species-typical attributes (typically aspects of wild behavior, biology or ecology: candidate *predictor* variables) with measures of species-typical welfare

(*e.g.* captive animals' infant mortality rates or behavioral problems: *outcome* variables). We summarize these studies next.

Focusing on captive Carnivora, Clubb and Mason (2003, 2007) collated data on stereotypic behavior across 33 species. After tests for serial independence to assess similarity between pairs of species (Abouheif, 1999), the authors used phylogenetic independent contrasts (PICs) to test two broad hypotheses. One was that wide-ranging species are at risk of stereotypic route-tracing (an idea proposed decades earlier by canid researchers [Forthman-Quick, 1984]); the other, that restricting hunting compromises well-being. Their analyses revealed that traveling large distances in the wild, and being both large-bodied and wide-ranging, were risk factors for route-tracing and elevated captive infant mortality (CIM). Reliance on hunting, in contrast, seemed not to predict poor captive welfare. The authors suggested that these results could inform collection planning: *"it might be sensible – both more cost-effective and humane – for zoos to focus on those carnivores inherently best suited to current, or at least readily achievable, enclosure sizes and enrichment/husbandry regimes. Wide-ranging species instead could be conserved in specialized breeding centres … or instead via in situ approaches"* (Clubb & Mason 2007). They also proposed that mimicking aspects of wide-ranging Carnivores' lives could enhance well-being via *"substantial increases in space; greater numbers of viewpoints; …more spatial and/or stimulus complexity and less day-to-day environmental predictability — combined (importantly) with the ability of the animals to control their own access to such increased variability; and more scope to approach or retreat from the public, conspecifics, and other stimuli, at will"*.

Capitalizing on new PIC software, a larger database, and an updated phylogeny, Kroshko *et al*. (2016) replicated this work. They confirmed the relationship between route-tracing and large daily travel distances/home range sizes, but found the latter no longer depended on body size, and that the daily travel distance effect was a mere by-product of home range size. This suggests the relationship between home range size and route-tracing is not mediated by active locomotion, leading the authors to re-emphasize the likely value of husbandry enhancements designed to emulate the variety and

control wide-ranging animals likely experience in the wild. Long chase distances also now tentatively emerged as a risk factor for route-tracing, albeit from a sample size of just five species. CIM, however, was no longer predicted by any aspect of wild biology, leaving the great variation in species-typical Carnivora CIM *"an urgent topic for future work, one best addressed using both a broader range of species-typical potential risk factors and [ZIMS] data on infant mortality"*. Partly to look at this, species differences in Carnivora welfare are now being re-investigated, incorporating six more years of data and applying a newer PCM approach, phylogenetic generalized least squares (PGLS) regressions (Bandeli, Mellor, & Mason, 2017; Mellor, Mendl, Bandeli, Cuthill, & Mason, 2017).

Next to use PCMs was a Swiss team who, with collaborators, investigated patterns in the mortality rates of zoo-housed Ruminants. Their main PCM study applied PGLS to 78 Ruminantia species (Müller *et al*., 2011). Each species' average captive life expectancy, expressed as a proportion of its maximum recorded life expectancy ('relative life expectancy': rLE) was used to assess husbandry success (Müller *et al*., 2011; Müller, Lackey, Streich, Hatt, & Clauss, 2010). Several hypotheses were tested about risk and protective factors for rLE, the authors arguing that results could help to optimize husbandry and identify types of species for which *"a higher husbandry success can more easily be achieved".* In later complementary studies, senescence rates (albeit combining non-PCM with PCM [PGLS] analyses) (Lemaître, Gaillard, Lackey, Clauss, & Müller, 2013) and seasonal mortality patterns (Carisch *et al.* 2017) were also examined for many of the same species.

One important hypothesis tested was that the intensive population management related to having a studbook would enhance rLE. This was supported: studbook-managed species had significantly longer rLEs than non-studbook species (Müller *et al.* 2011). Turning to intrinsic aspects of biology, two plausible hypotheses were rejected. One rejected hypothesis was related to natural social structure: that *"…density-dependent influences on LE (social stress, contact with pathogens) should have a higher impact in solitary and pair-living species, which are less adapted to crowded conditions (as in zoos)"*. Their second non-supported hypothesis was that, because most relevant zoos

were located in temperate regions, tropical species would have reduced rLE. Results instead "*indicat[ed] that climatic stress in (sub-)tropic species that are kept in the temperate zone does … not play an important role".* Subsequently, Carisch *et al.* (2017) similarly found that the latitude of a Ruminant species' origin (across 88 species) did not appear to predict over-winter mortality rates in zoos.

The fourth hypothesis tested by Müller *et al*. (2011) *was* supported. Mating system affected rLE: males from polygynous species were found to have reduced rLEs. Carisch *et al*. (2017) similarly found some apparent effects of mating system on seasonal mortality at the onset of rut in zoo-housed cervids of both sexes, leading them to advise that *"husbandry measures aimed at protecting females from rutting males are important, especially in cervids".*

Finally, a fifth, dietary-related hypothesis was also supported, inspired by a previous non-PCM result from deer suggesting that grazing species have longer rLEs in zoos than browsing species (Müller *et al*., 2010). In Müller *et al*. (2011) the same pattern emerged for female Ruminants: naturally grazing species were longer-lived in captivity, thus corroborating "*the subjective experience that browsers demonstrate a higher nutrition-related mortality in captivity and are more challenging to keep when compared with grazing species, owing to the complex logistics of providing browse".* Lemaître *et al.* (2013) subsequently analyzed senescence rates in a subset of 22 species for which age-specific wild mortality data were also available. Typically, aging rates were lower in zoos than in the wild, but this difference was most marked for grazers (of both sexes). These authors concluded *"this indicates that animals in zoos perform the better compared to free-ranging conditions the more they are grazers",* and emphasized again *"the difficulty of keeping browser species in captivity".*

Our final examples come from studies on Primates and the sole avian PCM study. Across 24 primate species, Pomerantz, Meiri, & Terkel (2013) used PGLS to reveal that, somewhat similar to the first Carnivora study (Clubb & Mason, 2003; 2007), long wild daily travel distances tended to predict stereotypic route-tracing. Additionally, large natural group sizes predicted another abnormal

behaviour, hair-pulling. These researchers argued that such research *"facilitates detection of the more 'susceptible' species, as well as enabling the decision-makers to focus on specific environmental factors in order to improve the primates' psychological welfare."* Specific husbandry recommendations were for activity levels to be increased (for instance via *"incorporating modular structures within the enclosure, allowing for easy and frequent change of the environment"*), along with the creation of *"more opportunities for positive social interactions for the animals. Where possible, it is recommended to house groups in numbers similar to those reported in the wild."*

Lastly, McDonald Kinkaid (2015) applied PICs to 201 Psittaciforme species kept by aviculturalists or as pets (work currently being replicated using PGLS and an updated phylogeny). Unlike the Primate study, she found no effect of sociality; however, naturally effortful modes of foraging and relatively large brains (a proxy for intelligence), were both risk factors for stereotypic behavior in captive pet parrots. Furthermore, naturally effortful modes of foraging also predicted reduced captive reproductive success, as did being classed as 'endangered', with a trend for similar effects of brain size. The author concluded: *"We can use this information to make informed predictions about the suitability for captivity of different species … my findings suggest that the two best predictors of this should be high natural foraging effort and large relative brain volume, such that species characterized by either one (or both) of these risk factors are intrinsically predisposed to adjust relatively poorly to captive conditions".* In terms of husbandry improvements, McDonald Kinkaid recommended supplying more naturalistic diets, and enrichment opportunities to learn and problem solve. She added, *"it would now be useful to perform comparative analyses for other similarly large-brained or relatively intelligent taxa – like corvids, primates, or cetaceans – in order to determine whether some of the same biological risk factors identified for parrots also predict relatively poor welfare among those groups".* This highlights some of the exciting research questions that PCMs are uniquely able to tackle (and we suggest more in the next section).

Using PCMs to test welfare-relevant hypotheses: a step-by-step guide

This section provides a 'how to' guide for future welfare studies, based on the studies just described, other relevant studies using zoo data, and PCM studies from other fields.

Hypothesis generation

As with any research, the first step is specifying the hypotheses and their predictions, since these determine which precise data are required (*e.g.* which specific species-level variables need quantifying). This might involve devising new hypotheses by reading about the species and welfare problems of interest, or instead identifying pre-existing hypotheses from the literature. Table 1 lists several published, but as yet untested, research ideas and their predictions: all ideal topics for future PCM studies.

Some specific hypotheses may arise from speculating about behavioral needs (like our 'does restricting flight affect bird welfare?' example earlier; or the question, 'do carnivores need to hunt?': see Table 1). Others may be inspired by patterns of unexplained variation in species-typical welfare. As we saw above, for instance, it was apparent differences between species that led to the testable hypotheses that being wide-ranging is a risk factor for route-tracing, and that being a browser predicts relatively short captive lifespans (with similar species differences generating the hypothesis that prey species hide their pain: see Table 1). In other cases, observed variation in captive welfare may prompt more open-ended research, with an array of competing hypotheses being tested. Why captive Carnivores show such variation in captive infant mortality is one case in point, potential predictors being any factor that could stress Carnivore mothers, and /or make infants more or less vulnerable to premature death.

TABLE 1 HERE PLEASE – UNTESTED HYPOTHESES

Data collection when using PCMs: general considerations

A single summary statistic is usually calculated, one per species, for each variable, and including as many species as possible maximizes power. Blomberg *et al.* (2003), for example, argue that at least 20 species are required for acceptable power and Type I error rates. Examples of species-level summary statistics might include (depending on the hypotheses under test): median home range size, whether or not a prey species, and/or median enclosure size (as potential predictor variables); and typical life expectancy, reproductive output, and /or median time budget spent on stereotypic behavior (as potential outcome variables). Where appropriate, median values are recommended over means, to reduce effects of outliers and skew in the raw data (Gittleman, 1989).

A key assumption of species-level summary statistics is that they do represent species-typical norms (Ives, Midford, & Garland, 2007): thus assuming that either intra-specific (*within* species) variation is absent (Garamszegi, 2014; Ives *et al.,* 2007), or the population has been sampled well enough that intra-specific variation is well-captured in the data (*e.g.* Garamszegi, 2014). However, intra-specific variability can sometimes be rather large (*e.g.* for behavioral traits), and biases can be introduced by, for instance, data collection differences, sub-population differences, and small and/or unequal sample sizes between species (Garamszegi & Møller, 2010). Practical ways to minimize such effects of intra-specific variation are to use data from as many individuals from as many locations as feasible; and/or to impose a minimum sample size of individuals per species (on determining suitable intra-specific sample sizes and for analyses overall see: Garamszegi, 2014). For example, in the Kroshko *et al.* (2016) and McDonald Kinkaid (2015) studies, for inclusion each species had to be represented by ≥5 individuals; while in Müller *et al.* (2010; 2011) the minimum was 45. Another possible solution, though not always feasible, is to statistically account for intra-specific variation and sampling error (Symonds & Blomberg, 2014), which then improves the estimation of parameters when models are run (Ives *et al.,* 2007) – an approach touched on further below.

Next we describe potential sources of and methods for collecting data on health and welfare outcome variables, outlining the benefits and limitations of each. After this, we turn to potential predictor variables.

Outcome variable: data collection

Accessing zoo-generated datasets, *e.g.* ZIMS and studbooks

CIM values in the Carnivora studies (Clubb & Mason, 2003, 2007; Kroshko *et al*., 2016) were extracted from *International Zoo Yearbooks*, which published infant mortality reports. This practice stopped in 2000, however, making this data source increasingly historical. In collaboration with individual zoos, veterinary records can also be useful sources of data relating to health and disease (*e.g.* Miller, Hogan, & Meehan, 2016). For longevity and mortality data, studbooks (in collaboration with species' studbook keepers) are also potentially useful, though not used in PCM research as yet. An alternative, up-to-date, and far more extensive source of data is the Zoological Information Management System (ZIMS), a worldwide member zoo database. Research requests can be made to "Species 360" (formerly the International Species Information System [ISIS]) for ZIMS husbandry- or veterinary-related data and studbook information on individuals from many thousands of species (Species360, 2018). ZIMS data can potentially provide impressive statistical power, and also widely samples global populations. For example, access to ZIMS data allowed Müller *et al*. (2011), to create a dataset of 166,901 individuals across 78 species for their Ruminantia study. However, if planning to use ZIMS data, one practical consideration is that applications for access can take up to a year to be processed, and may be denied (*e.g.* Kroshko, 2015; McDonald Kinkaid, 2015). For relatively short-term studies, such as PhD projects, relying on such data is therefore unwise.

Whatever the source, data must always be checked for errors. In ZIMS, for example, husbandry and veterinary data quality and type can vary between collections; birth and death records

can be missing; and animals may be lost to audit when moved between institutions (*e.g.* Mace & Pelletier, 2007; L. Rowden, 2015 pers. comm.). Therefore, data should be checked for obvious errors and internal consistency, and ideally cross-referenced with other sources and/or validated by contacting zoos holding individuals with dubious entries (*e.g.* Clubb *et al.,* 2009). Questionable data that cannot be corrected by such means should be excluded.

Surveying animal carers, *e.g.* keeping staff

Surveys can cheaply, efficiently gather data on many animals from many collections (*cf.* Lewis, Shepherdson, Owens, & Keele, 2010; Munson, 1993): ideal for PCM studies. For example, using an online survey of pet owners, McDonald Kinkaid (2015) quite rapidly collected data on stereotypic behavior and medical conditions for over 1,400 individual pet parrots from 74 species worldwide. Surveys can also capture a large sample of the overall population, making them good for assessing prevalence (% affected animals), since this measure only requires simple yes/no answers from respondents. However, time demands on animal care staff must be considered: zoo surveys, for instance, should be designed thoughtfully to not over-burden staff time and enhance chances of good return rates (see: Plowman, Hosey, & Stevenson, 2006). Surveys are less effective for quantitative data like behavioral time-budgets, since it is unrealistic to expect participants to make lengthy behavioral observations. Noise is also likely to affect survey data due to idiosyncratic responses from different people, although this can be mitigated by surveying numerous species and building up large sample sizes for each one.

Extracting data from published research

Using ready-published data is efficient, and can also allow coverage of diverse collections worldwide. The Carnivora studies extracted observational data on stereotypic behavior from 173 studies using this method (Clubb & Mason, 2003, 2007; Kroshko *et al*., 2016), yielding data on over 1,300 individuals across 51 species from collections worldwide. For each species, values were summarized across individuals to calculate a single statistic: median % observations spent

stereotyping by affected animals. McDonald Kinkaid (2015) likewise used publications to obtain Psittaciforme reproduction data: captive hatch rates for 122 species from North America and, using an expert technical report, the relative breeding difficulty of 141 species.

This approach has limitations, however. If combining multiple sources, data will not be standardized (*cf.* Garamszegi & Møller, 2010), necessitating quality checks. For instance, for behavioral data, included studies should use consistent data collection methods, and arguably focus on stably housed subjects (since recent changes in *e.g.,* social grouping or enrichment, may affect how representative behavioral data are). The resulting dataset will likely contain gaps, with some species being well-represented but others under-sampled or even absent, sometimes because of study biases (*cf.* Melfi, 2009). Furthermore, in Carnivora (and perhaps other taxa), research is skewed towards stereotypic over non-stereotypic individuals, especially in enrichment studies (Kroshko, 2015). This biased sampling means that subjects are not a random sample of the overall population, preventing accurate estimates of prevalence and/or true population means.

Collecting data by direct observation

Direct observation is ideal for collecting accurate, standardized behavioral data. Furthermore, for stereotypic behavior, both prevalence and average time budgets can be calculated, since populations can be sampled at random, without biases towards stereotypers. But this data quality comes at a price: direct observation is time-consuming and financially costly, potentially limiting a study's scope. Thus Pomerantz *et al.* (2013) only assessed 214 individuals (albeit representing 24 species), from just six Israeli zoos. This may introduce noise from individual and site idiosyncrasies and limit their findings' generalizability. A potential solution to this problem, if applied with care, might be to use zoo visitors (*cf. e.g.* Williams, Porter, Hart, & Goodenough, 2012) to collect data across multiple sites and even countries: an approach not yet used in PCM research.

Outcome variable: sources of potential confounds

As mentioned above, outcome variable data can be affected by several sources of noise (non-systematic error), and by biases towards certain species and individuals. Two potential sources of systematic confound can further influence welfare-related variables: extrinsic effects of captive management, and intrinsic differences in pace of life and reproductive strategy.

Extrinsic influences of captive management can clearly affect outcome variables (illustrated in the studbook effects on Ruminant lifespan, for example). In some cases this might add noise to data. For example, Müller *et al*. (2010; 2011) recognised that a potential confound specific to their Ruminata studies was the selective culling of surplus animals which, if unaccounted for, would artificially reduce species' rLEs. This was handled during data processing by excluding animals that died within two years of birth (so potentially culled). In other cases, management effects may, if they vary *systematically* across species, potentially create artefactual relationships between intrinsic species-typical attributes and outcome variables (thus Type I errors), or mask real relationships (causing Type II errors). To illustrate with Kroshko *et al*. (2016)'s Carnivora study, typical housing conditions were found to covary with annual home range size: naturally widely ranging species were often kept in enclosures with little cover. How can one ensure that the home range effect on route-tracing really results from home range size and not this correlated aspect of husbandry? The answer is to include such confounding variables as statistical controls (as these authors did: home range size really does predict route-tracing, even after controlling for the amount of cover: Kroshko *et al*., 2016).

Considering extrinsic effects of captive management during early stages of the research allows appropriate husbandry and environmental data to be collected, or gleaned from publications or survey questions. One way to then check for potential confounds (*cf.* Kroshko *et al*., 2016) is to correlate each husbandry/environmental variable against each species-typical wild attribute (using appropriate PCM tests). Should significant relationships emerge, such confounds can be controlled for by inclusion in subsequent models. This method is particularly useful for datasets with missing values for different variables. Another approach (*cf.* Pomerantz *et al*., 2013) is to include husbandry variables

into all hypothesis-testing models (though unless the dataset is complete, this will cause the loss of species with incomplete data).

The second major potential source of confound is intrinsic variation in pace of life and reproductive strategy. For example, simply using infant mortality (IM) or reproductive rate as welfare indicators would be naïve since these can reflect intrinsic, evolved differences between species in reproductive strategy (*e.g.* whether infants are altricial). Using maximum recorded captive lifespan as an outcome variable would also be inappropriate, as lifespan is intrinsically related to body mass (smaller-bodied species tending to have shorter lives than larger species, *sensu* Hill, 1950). Therefore, for life history-related outcome variables, evolved intrinsic differences between species must be understood and factored in before inferring any effects of captivity. For example, when Clubb and Mason (2003, 2007) found that natural home range sizes predicted CIM, they then gained data on wild IMs to check that wide rangers did not just naturally have high IMs. Similarly, McDonald Kinkaid (2015) corrected parrots' hatch rates in captivity with wild rates, to control for intrinsic species differences in reproductive rate. Finally, in Müller *et al.* (2010) and Müller *et al.* (2011)'s Ruminant work, as we saw, intrinsic life expectancy was controlled for by calculating the ratio of mean lifespan in captivity to the maximum ever recorded for each species. Response to captivity *per se* could then validly be inferred from this derived 'rLE' outcome variable.

Predictor variables: data collection

All the PCM studies described above used published sources (*e.g.* journals) to obtain values for species-typical values for potential predictor variables: a cost-effective, although quite time-consuming, method. Just as for outcome variables, calculated species-typical values are likely to be more accurate if gleaned from many sources; and researchers should ensure these meet quality criteria based on, *e.g.,* techniques used, representativeness of the wild populations sampled, data collection

time periods, *etc.* (see *e.g.* Clubb & Mason, 2007). Research effort (estimated based on the number of papers published per species) can also be controlled for by including it in statistical models if a potential confound (*e.g.* for estimates of species-typical innovation rates, see McDonald Kinkaid 2015, following Overington, Morand-Ferron, Boogert, & Lefebvre, 2009). Once quality data have been compiled, values can then be appropriately summarized to yield a single summary value per species. Note that it is most efficient to collect predictor variable data after the subset of species with good quality outcome data has been identified. However, the collation of predictor variable data should ideally be conducted blind to outcome variable values, to avoid risks of bias.

Textbooks can also yield values for species-typical attributes (*cf.* McDonald Kinkaid, 2015; Müller *et al*., 2011), as can books and theses from specialist university libraries. Experts can also be valid sources: they may have unpublished information or be able to provide estimates of species-typical attributes. Furthermore, for some taxa there are freely available databases collating species-typical wild ecology and behavior from many studies, *e.g.* Mammalian Species accounts (American Society of Mammalogists, 2017) and PanTHERIA (Jones *et al.,* 2009) (see also Ecological Archives: ESA, 2016). Both Kroshko *et al.* (2016) and Pomerantz *et al.* (2013) took advantage of such sources. Using these databases is highly time-saving, although inclusion criteria and sources used should be carefully assessed to judge data quality (Bielby *et al.,* 2007; Lemaître, Müller, & Clauss, 2014).

In some instances, researchers may be unable to find the precise predictor variable data needed to test a hypothesis. For example, despite flight and travel distances being plausible potential risk factors for poor parrot welfare, McDonald Kinkaid (2015) could not find these data. Quantitative data on daily foraging activity budgets of wild Psittaciformes were also scarce. However, based on characteristics of the main food in typical wild diets, and consulting with experts, McDonald Kinkaid (2015) devised a simple, broad categorical descriptor: relatively 'high' or 'low' natural foraging effort, allowing her to investigate relationships between relatively effortful natural foraging and captive welfare. Generating valid predictor variables may thus require some lateral thinking. Finally, while data on species-typical behavior and biology are the typical predictors, one study used the

discrepancies between wild and captive norms. Pomerantz *et al.* (2013) collected data on average primate group sizes in zoos and those in the wild to create a 'group size ratio' (captive/wild) predictor. This quantified the degree of mismatch between wild and captive conditions for each species; the greater the mismatch (*i.e.* smaller values), the more hair-pulling was observed (see Fig. 2 in Pomerantz *et al.,* 2013). This could be a useful approach for future welfare PCM researchers to consider.

Predictor variables: sources of potential confounds

Correlated aspects of species' biology are the main sources of confound for predictor variables. One relevant example is body size, which co-varies with many aspects of most species' life-histories and biology (*e.g.* Gittleman, 1986). In Carnivora, for example, body size co-varied with home range size (larger-bodied species having larger home ranges). Had analyses naïvely been performed to test the predictive power of home range size alone, body size *per se* could have explained the apparent range size effect. Here, this was managed by including body size in home range size models as a covariate, allowing assessment of home range size effects independently of body size (Clubb & Mason, 2003, 2007; Kroshko *et al*., 2016). Another example is daily distance traveled and home range size, which both emerged as correlated risk factors for route-tracing. Kroshko *et al.* (2016) disentangled their effects by including both predictors into the same model, so revealing the apparent daily distance traveled effect to be merely a 'side-effect' of home range size. Ongoing research is now investigating further whether other natural correlates of Carnivora home range (such as metabolic rate) are the true predictors of route-tracing (Bandeli *et al.,* 2017).

To avoid correlated aspects of species' biology acting as confounds, reading about your species of interest is thus essential, as this pre-warns of interrelated aspects of species' biology. Checking for collinearity between predictor variables within your dataset is also good practice, including any correlates as covariates in final models where appropriate. Another solution might be to run so-called "phylogenetic path analyses" (Gonzalez-Voyer & Von Hardenberg, 2014; Hardenberg & Gonzalez-Voyer, 2013; van der Bijl, 2017), a topic we outline in Table 2.

Data analysis and interpretation

Creating the dataset

Microsoft Excel is a good software package for collating data, calculating species' summary statistics, and constructing final comparative datasets (with summary statistics for outcome and predictor variable(s) arranged in columns, and each species in its own row). Microsoft Access is also useful (though more challenging to use) for constructing complex databases, especially for parsing out different types of data from various sources. From these spreadsheets or databases, data can then easily be transferred into statistical packages such as R, Mesquite, *etc.* (see Table 2). Careful data entry checks for errors and outliers should be made throughout all calculations and the final dataset construction.

Sourcing phylogenetic trees

All PCM studies require a phylogenetic 'tree' for the species of interest (*e.g.* Cornwell & Nakagawa, 2017). Effectively a branching diagram depicting hypothesized relationships among species (Baum, 2008) (see Fig. 2), this sums up how closely related – and thus potentially non-independent – the species are. For analyses, trees usually need to be in NEXUS or Newick formats, which are readable for PCM software (Felsenstein *et al.,* 1990; Maddison, Swofford, & Maddison, 1997). For most taxa, tree files can be sourced by searching ecological and evolutionary literature, and generally speaking, newer trees are preferable since they reflect the most up-to-date knowledge on relationships between taxa (Arnold *et al.,* 2010). The structure of the tree, and thus relatedness of the species, is then factored in during PCM analyses (as discussed below).

****FIGURE 2 HERE PLEASE: TREE***

Trees only display *hypothesized* relationships between species: when reconstructing historic events, some uncertainty surrounds the precise patterns and/or timings of phylogenetic relationships

(*sensu* Arnold *et al*., 2010). Consequently, generally no one tree is definitively correct, and trees are constantly being refined to reflect updated knowledge. Nevertheless, PCM statistical tests assume that a given tree's topology and relationships between species *are* known and correct (Felsenstein, 1985; Garland, Bennett, & Rezende, 2005). Therefore, evolutionary researchers often use consensus trees that sum up agreement between multiple trees (Adams, 1972). Also it is recommended to perform analyses across a tree 'block' (a set of similar, though slightly different, likely trees: *e.g.* Arnold *et al*., 2010). More certain nodes appear more frequently in the block, less certain nodes less frequently, representing uncertainty in the phylogeny (Arnold *et al*., 2010). Analyses are then performed across the whole block (with this accounting for topological and branch length uncertainty), so producing robust results for which associated confidence intervals can also be generated (*e.g.* Arnold *et al*., 2010; Jetz, Thomas, Joy, Hartmann, & Mooers, 2012). Results are then reported, not only as summary values (*e.g.* median *P* values and slopes) but also with the 95% confidence intervals for each (*e.g.* Su, Cassey, Vall-llosera, & Blackburn, 2015). Tree blocks are usually freely available from online literature (*e.g.* Arnold *et al*., 2010; Jetz *et al*., 2012; Jetz *et al*., 2014; TreeBASE, 2016), with particularly useful exercise files and data from: www.10ktrees.nunn-lab.org/howToUse.html.

Running the statistical analyses

As already outlined, two commonly used PCM statistical tests are PICs and PGLS. Both are extensively reviewed elsewhere, so here we provide brief outlines, highlighting key papers for further reading. Other types of PCM test are also available, depending on specific requirements (Table 2 summarizing some of these).

Phylogenetic independent contrasts (PICs)

The rationale underlying PICs (Felsenstein 1985), is that while related species are non-independent, the *differences* – or contrasts – between them *are* independent, representing evolution since the species diverged. PICs assume that more recently diverged species (typically those with shorter branch lengths since their last shared node) will be particularly similar to each other, because

little time has passed since divergence. In contrast, species diverging long ago (with relatively long branch lengths since their last shared node) are assumed to be less similar to each other, because more time has passed, and more evolution occurred, since divergence. This assumption, which PICs are robust to minor deviations from (Diaz-Uriarte & Garland, 1996), reflects the so-called 'Brownian Motion' model of evolution: one of genetic drift with no selection, simply occurring to a greater extent with the passage of more evolutionary time (*sensu* Felsenstein, 1985).

PICs handle this as follows. First, for each variable, contrasts are calculated between pairs of species or ancestral nodes. Thus the original set of N non-independent species datapoints are converted to $N - 1$ independent contrasts (see Fig. 3). Next, each contrast is divided by its standard deviation (the square root of the sum of the relevant branch lengths), to account for how much evolutionary time has passed since divergence (Felsenstein, 1985). These 'standardized contrasts' are now suitable for conventional statistical analysis (*e.g.* Gittleman & Luh, 1992). Note that because the absolute distance of each contrast from the origin is important, rather than their positions relative to one another, regressions using contrasts should be forced through the origin during analyses (*e.g.* Garland, Harvey, & Ives, 1992).

Felsenstein (2008) has more recently extended his original PIC method to account for the intra-specific variation discussed earlier (*e.g.* that caused by sampling error). Values from individuals are used to calculate species-specific means for each attribute, and a weighting factor based on each species' sample size is then incorporated into the contrasts calculations.

FIGURE 3 HERE PLEASE (PIC FIGURE)

Phylogenetic generalized least squares regressions (PGLS)

PGLS (Grafen, 1989) involves linear regression models that effectively incorporate the tree's topology and branch lengths into the regression equation (Freckleton, Harvey, & Pagel, 2002; Garland *et al*., 2005; Pagel, 1999). During PGLS, the model estimates how similar species' trait values would

be if they evolved by Brownian Motion. The earliest forms of PGLS then incorporated statistical controls for this pattern into each analysis, to 'partial it out' and so prevent it from influencing the final results. In later refinements of PGLS, models assess the extent to which this pattern (i.e. the one expected if traits evolved by Brownian Motion) actually occurs in the dataset being analyzed, in turn allowing this degree of phylogenetic signal to be statistically corrected for. This highlights a major benefit of modern-day PGLS: the generation of a metric termed 'Pagel's Lambda' that captures the degree of phylogenetic signal present (Freckleton *et al*., 2002; Pagel, 1999; Revell, 2010). Lambda varies from 1 (strong signal, as implied in Brownian Motion) to 0 (no signal, with even close sister species being statistically independent) (Pagel, 1999). When lambda is 1, PGLS thus returns results identical to PICs; when lambda is 0, PGLS performs near identically to standard regressions which treat species' datapoints as independent (Pagel, 1999; Revell, 2010); while at intermediate values, the non-independence between species is corrected for according to the amount of inter-correlation found (Pagel, 1999; Revell, 2010). Thus, PGLS can flexibly control for the actual amount of signal present rather than, as with PICs, assuming strong signal *is* present.

PGLS can also potentially incorporate different evolutionary models (*e.g.* by using the R Package 'ape': Paradis, 2012). These include: Ornstein-Uhlenbeck, which models 'stabilizing selection', wherein attribute evolution is constrained within an optimum range (more realistic for some attributes, *e.g.* mammalian body size [Harmon *et al*., 2010]); and other evolutionary patterns, like adaptive radiation, where attributes evolve rapidly immediately after species diverge, slowing towards the tips of the tree (Early Burst: Harmon *et al*., 2010). Furthermore, intra-specific variation can potentially be incorporated into PGLS models, by providing the standard error associated with attributes' values, or inputting individuals' values for each attribute (Revell, 2012 based on Ives *et al*., 2007). For further reading, Symonds and Blomberg (2014) provide an excellent overview of PGLS.

***TABLE 2 HERE PLEASE - DIFFERENT TESTS ***

Presenting and interpreting the results

The results of PCM analyses – test statistics, *P* values, degrees of freedom, and effect sizes like R^2 – are reported in much the same way as those for other statistical outputs. For PGLS, Pagel's Lambda (Pagel, 1999) should also be reported. Often effect sizes can be rather small, even when models and terms are significant (Freckleton, 2009). Effect sizes should therefore be reported to permit appropriate interpretations of results. Conversely, when sample sizes are small ($N = <20$, *sensu* Blomberg *et al*., 2003), non-significant results may represent low statistical power rather than truly absent relationships. Additionally, if analyses are performed across a tree block, then results can and should include the associated 95% confidence intervals (which indicate the ranges within which true values likely lie).

Visual representations, such as graphs of correlations between attributes, are also useful. They can provide visual information, not just on effect sizes and the explanatory value of predictor variables, but also on potential outliers, and on thresholds that may exist (below which species have no apparent welfare problems, but above which signs of poor welfare are clear). However, because graphs from PIC models use contrast values rather than species' values, axes values will be somewhat arbitrary and non-informative (despite accurately depicting relationships). Alternatively, species' values may be plotted on the relevant axes as *per* standard regressions. This will not illustrate the PCM statistics run (as should be stated in figure legends), but is more intuitive to understand. Researchers using PGLS should ensure they plot the PGLS regression line, which will be 'weighted' appropriately to that particular analysis (Symonds & Blomberg, 2014).

By validly testing hypotheses to identify risk and protective factors, PCM results can be interpreted in three main ways to improve animal welfare. First, they can identify vulnerable 'problem' species to target for special care. Second, the principles they yield can help predict which additional species beyond those in the dataset, are likely, or not, to be at risk of health and welfare problems. Third, they can generate novel ideas about how to improve husbandry (*e.g.* via altering species' typical diets or enclosure characteristics). Such manipulations can also test the causality of

relationships. PCM outputs merely represent *correlations* between wild biology predictors and welfare-relevant outcomes: even after carefully considering the various correlates of predictor variables and systematic confounds of outcome variables, and performing 'path' analyses (see Table 2), true causality cannot be assumed (*sensu* Garland *et al*., 2005; Gittleman, 1989). But PCM outputs do yield causal *hypotheses,* which can potentially be tested experimentally. PCMs can thus inform future work that uses the other two complementary approaches to welfare research, the results of which can help to both test causal hypotheses and improve animal husbandry.

Discussion

We hope this review encourages more researchers to use PCMs to investigate health and welfare issues in captive wild animals. Many members of wild and semi-wild species are housed in contexts as diverse as farms and domestic homes (Mason *et al.,* 2013), and this approach could be relevant to all of them if applied by aquaculturalists, aviculturalists, and others. However, PCMs are particularly useful for zoos and aquaria for the following reasons: these institutions actively aim to promote animal welfare in evidence-based ways; much research has already been conducted on zoo animals, making existing theses and publications ripe for collation and meta-analysis; zoos have already collected vast amounts of welfare-relevant data (*e.g.* via ZIMS), the incredible value of which has barely been harnessed; and finally, the size and diversity of global zoological collections (Conde, Flesness, Colchero, Jones, & Scheuerlein, 2011) makes the "20 species" minimum suggested by some (*e.g.* Blomberg *et al.,* 2003) an easy challenge to meet. Zoo researchers are thus enviably positioned to capitalize on PCM-based research. We recognize that applying PCMs can be technically daunting: good trees must be sought and potentially complex statistical analyses performed. But many helpful resources are readily available: papers and books (we particularly recommend Arnold *et al.,* 2010; Freckleton, 2009; and Symonds & Blomberg, 2014); online lecture notes, primers, and forums; software support; and experts in many universities' biology, ecology, and zoology departments.

Furthermore, the benefits of running PCMs to investigate welfare issues are worth the effort.

For one, PCMs represent an economical, efficient complement to experimentation and epidemiology – the traditional methods for studying zoo animal health and welfare. Using the published literature alone, we found nearly 20 untested hypotheses that are ideal for testing using PCMs (Table 1), with many more possible beyond these. Furthermore, PCMs can address welfare research questions that would be logistically or ethically impossible to investigate in other ways (*cf.* Clubb & Mason, 2004; Clubb & Mason, 2007). Such questions include whether being intelligent, or unable to hunt, migrate, or fly, compromises welfare in captive animals, and whether prey animals have evolved to mask states of pain or sickness from potential predators. PCMs are unique in making these fundamentally fascinating and practically important questions amenable to study.

Second, just like other methods of investigating welfare issues, PCM results can indicate effective ways in which to improve husbandry; but because they can also address questions hard to tackle via experimentation or epidemiology, the insights they yield can be novel. As reviewed above, PCMs have thus generated evidence-based recommendations to supply Carnivores with more variety and control, in order to reduce route-tracing; to provide Ruminants that browse in the wild with more natural diets in captivity, in order to reduce nutrition-related mortality; to protect female Ruminants, especially Cervidae, from rutting males in order to reduce seasonal deaths; to house zoo primates in naturalistic group sizes, and encourage active travel, in order to reduce two forms of abnormal behaviour; and to supply captive parrots with more naturalistic diets, along with opportunities to learn and problem-solve, in order to enhance their welfare. Furthermore, by identifying the most 'susceptible' species (*cf.* Pomerantz *et al.,* 2013), such as Ruminant species with the lowest rLEs (Müller *et al*., 2011), PCMs can highlight those it might be most important to target for improvements.

A third major benefit of using PCMs to investigate welfare issues is that the data collated and results generated could have great benefits for global collection planning. PCMs, as we have seen, can identify both specific species intrinsically unlikely to thrive in zoo conditions, and also broad *types* of

species at risk of welfare problems. While one response may be to target such species for special care, an alternative is to phase them out in favor of species revealed to be intrinsically likely to thrive in zoo conditions and protected from welfare problems. Such recommendations may be negatively received by some in the zoo community, but we propose that factoring animal welfare into collection planning is both strategic and practical (see Table 3). Most of the 4000 species currently represented in zoos are in populations too small to be viable (*e.g.* Conde *et al.,* 2013; Lees & Wilcken, 2009). Furthermore, zoos have limited spatial and financial resources (*e.g.* Conde *et al.,* 2013; Fa, Funk, & O'Connell, 2011; Gusset, Fa, & Sutherland, 2014; Lees & Wilcken, 2009; McGowan, Traylor-Holzer, & Leus, 2017). It is thus important to decide which species zoos should focus on (*e.g.* Conde *et al.,* 2013; Fa et *al.,* 2011; Gusset *et al.,* 2014). Endangeredness has been proposed as one criterion, but in reality, diverse factors are currently at play in collection planning (Bowkett, 2014; Fa *et al*., 2011). If species-typical welfare explicitly played a role in such decisions, the benefits would range from practical and economic advantages to improvements in the viability and conservation relevance of captive populations (*e.g.* due to reduced rates of domestication) (see Table 3). PCMs could thus be part of a holistic approach involving evidence-based, welfare-guided collection planning, with the ultimate goal of ensuring that all captive populations can readily be kept successfully (*e.g.* Alroy, 2015; Conde *et al.,* 2013; Gusset *et al.,* 2014; Kaumanns & Singh, 2015).

TABLE 3 HERE PLEASE: BENEFITS OF CONSIDERING WELFARE

Conclusions

1. PCMs are powerful tools for captive wild animal welfare researchers, yielding novel insights into welfare that can be both practically and fundamentally valuable.

2. Results of PCM studies can be used to improve welfare, by guiding tailored changes to husbandry and enrichment; and to help prioritize conservation efforts, by identifying and even *predicting* which species can be most viably cared for practically, financially, and ethically.
3. The staggering array of species held in zoos means they have potential to be a phenomenal research resource. With PCMs used in relatively few welfare studies to date, and many untested hypotheses still awaiting research, there is now plentiful scope for zoo researchers and others to further develop and apply this exciting methodology.

Acknowledgments

We would like to thank Sally Sherwen, Mark Puttick, Miranda Bandeli, María Díez-León, Jamie Ahloy Dallaire, Sean Nee, Hafiz Maherali, Innes Cuthill, Jason Watters and a referee for their useful thoughts and comments on earlier versions of this manuscript. We would also like to thank Mike Mendl for helpful discussion on some of the topics covered in this review.

References

Abouheif, E. (1999). A method for testing the assumption of phylogenetic independence in comparative data. *Evolutionary Ecology Research, 1*, 895-909.

Adams III, E. N. (1972). Consensus techniques and the comparison of taxonomic trees. *Systematic Biology, 21*(4), 390-397.

Alroy, J. (2015). Limits to captive breeding of mammals in zoos. *Conservation Biology, 29*, 926-931.

Appleby, M. C., Olsson, A., & Galindo, F. (2018). *Animal Welfare*: CABI.

Arnold, C., Matthews, L. J., & Nunn, C. L. (2010). The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews, 19*, 114-118. doi:10.1002/evan.20251

AZA. (2018). Animal Welfare Committee. Retrieved from https://www.aza.org/animal_welfare_committee

Bandeli, M., Mellor, E., & Mason, G. (2017). *Why does home range size predict captive Carnivora welfare?* Paper presented at the 3rd International Symposium on Zoo Animal Welfare, Chicago, USA.

Baum, D. (2008). Reading a phylogenetic tree: The meaning of monophyletic groups. *Nature Education, 1*, 190.

Bielby, J., Mace, G. M., Bininda-Emonds, O. R. P., Cardillo, M., Gittleman, J. L., Jones, K. E., Orme, D. L. C., Purvis, A. (2007). The Fast-Slow Continuum in Mammalian Life History: An Empirical Reevaluation. *The American Naturalist, 169*, 748-757. doi:10.1086/516847

Björklund, M. (1997). Are 'Comparative Methods' Always Necessary? *Oikos, 80*, 607-612. doi:10.2307/3546637

Blay, N., & Côté, I. M. (2001). Optimal conditions for breeding of captive humboldt penguins (*Spheniscus humboldti*): A survey of British zoos. *Zoo Biology, 20*, 545-555. doi:10.1002/zoo.10002

Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution, 57*, 717-745. doi:10.1111/j.0014-3820.2003.tb00285.x

Bowkett, A. E. (2014). Ex situ conservation planning is more complicated than prioritizing the keeping of threatened species in zoos. *Animal Conservation, 17*, 101-103. doi:doi:10.1111/acv.12116

Britt, S., Cowlard, K., Baker, K., & Plowman, A. B. (2015). Aggression and self-directed behaviour of captive lemurs (*Lemur catta*, *Varecia variegata*, *V. rubra* and *Eulemur coronatus*) is reduced by feeding fruit-free diets. *Journal of Zoo and Aquarium Research, 3*, 52-58.

Bronson, E., Bush, M., Viner, T., Murray, S., Wisely, S. M., & Deem, S. L. (2007). Mortality of captive black-footed ferrets (*Mustela nigripes*) at Smithsonian's National Zoological Park, 1989-2004. *Journal of Zoo and Wildlife Medicine, 38*, 169-176. doi:10.1638/1042-7260(2007)038[0169:MOCBFM]2.0.CO;2

Byrne, P. G., & Silla, A. J. (2017). Testing the effect of dietary carotenoids on larval survival, growth and development in the critically endangered southern corroboree frog. *Zoo Biology, 36*, 161-169. doi:10.1002/zoo.21352

Cardillo, M., Mace, G. M., Jones, K. E., Bielby, J., Bininda-Emonds, O. R. P., Sechrest, W., Orme, D. L. C., Purvis, A. (2005). Multiple Causes of High Extinction Risk in Large Mammal Species. *Science, 309*, 1239-1241. doi:10.1126/science.1116030

Carisch, L., Müller, D. W. H., Hatt, J. M., Lackey, L. B., Rensch, E. E., Clauss, M., & Zerbe, P. (2017). Seasonal mortality in zoo ruminants. *Zoo Biology, 36*(1), 74-86. doi:10.1002/zoo.21337

Clubb, R., & Mason, G. (2004). Pacing polar bears and stoical sheep: testing ecological and evolutionary hypotheses about animal welfare. *Animal Welfare, 13*, 33-40.

Clubb, R., & Mason, G. J. (2007). Natural behavioural biology as a risk factor in carnivore welfare: How analysing species differences could help zoos improve enclosures. *Applied Animal Behaviour Science, 102*, 303-328. doi:10.1016/j.applanim.2006.05.033

Clubb, R., Rowcliffe, M., Lee, P., Mar, K. U., Moss, C., & Mason, G. J. (2009). Fecundity and population viability in female zoo elephants: problems and possible solutions. *Animal Welfare, 18*, 237-247.

Clutton-Brock, T. H., & Harvey, P. H. (1977). Primate ecology and social organization. *Journal of Zoology, 183*, 1-39. doi:10.1111/j.1469-7998.1977.tb04171.x

Conde, D. A., Colchero, F., Gusset, M., Pearce-Kelly, P., Byers, O., Flesness, N., Browne, R. K., Jones, O. R. (2013). Zoos through the lens of the IUCN Red List: a global metapopulation approach to support conservation breeding programs. *Plos One, 8*, e80311.

Conde, D. A., Flesness, N., Colchero, F., Jones, O. R., & Scheuerlein, A. (2011). An Emerging Role of Zoos to Conserve Biodiversity. *Science, 331*, 1390-1391. doi:10.1126/science.1200674

Cornwell, W., & Nakagawa, S. (2017). Phylogenetic comparative methods. *Current Biology, 27*, R333-R336. doi:https://doi.org/10.1016/j.cub.2017.03.049

Couquiaud, L. (2005). Whales, Dolphins, and Porpoises: Presentation of the Cetaceans. *Aquatic Mammals, 31*, 288-310.

Cuthill, I. C. (2005). The study of function in behavioural ecology. *Animal Biology, 55*, 399-417.

Diaz-Uriarte, R., & Garland, T. (1996). Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Systematic Biology, 45*, 27-47.

Díez-León, M., Bowman, J., Bursian, S., Filion, H., Galicia, D., Kanefsky, J., Napolitano, A., Palme, R., Schulte-Hostedde, A., Scribner, K. (2013). Environmentally enriched male mink gain more copulations than stereotypic, barren-reared competitors. *Plos One, 8*, e80494.

Diniz-Filho, J. A. F., & Torres, N. M. (2002). Phylogenetic comparative methods and the geographic range size – body size relationship in new world terrestrial carnivora. *Evolutionary Ecology, 16*, 351-367.

Ducatez, S., Clavel, J., & Lefebvre, L. (2015). Ecological generalism and behavioural innovation in birds: technical intelligence or the simple incorporation of new foods? *Journal of Animal Ecology, 84*, 79-89. doi:10.1111/1365-2656.12255

EAZA. (2013). *The Modern Zoo: Foundations for Management and Development*. EAZA Executive Office, Amsterdam.

ESA. (2016). Ecological Archives. Retrieved from http://esapubs.org/archive/default.htm

Fa, J. E., Funk, S. M., & O'Connell, D. (2011). *Zoo conservation biology*: Cambridge University Press.

Felsenstein, J. (1985). Phylogenies and the Comparative Method. *The American Naturalist, 125*, 1-15.

Felsenstein, J. (2008). Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *The American Naturalist, 171*, 713-725.

Felsenstein, J. (2016). PHYLIP. Retrieved from http://evolution.genetics.washington.edu/phylip.html

Felsenstein, J., Archie, J., Day, W. H. E., Maddison, W., Meacham, C., Rohlf, F., & Swofford, D. (1990). The Newick tree format. Retrieved from http://evolution.genetics.washington.edu/phylip/newicktree.html

Fisher, D. O., & Owens, I. P. F. (2004). The comparative method in conservation biology. *Trends in Ecology & Evolution, 19*, 391-398. doi:http://dx.doi.org/10.1016/j.tree.2004.05.004

Freckleton, R. P. (2009). The seven deadly sins of comparative analysis. *Journal of Evolutionary Biology, 22*, 1367-1375. doi:10.1111/j.1420-9101.2009.01757.x

Freckleton, R. P. (2011). Dealing with collinearity in behavioural and ecological data: model averaging and the problems of measurement error. *Behavioral Ecology and Sociobiology, 65*, 91-101. doi:10.1007/s00265-010-1045-6

Freckleton, R. P. (2012). Fast likelihood calculations for comparative analyses. *Methods in Ecology and Evolution, 3*, 940-947.

Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic Analysis and Comparative Data: A Test and Review of Evidence. *The American Naturalist, 160*, 712-726. doi:10.1086/343873

Garamszegi, L. Z. (2014). Uncertainties due to within-species variation in comparative studies: measurement errors and statistical weights *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 157-199): Springer.

Garamszegi, L. Z., & Møller, A. P. (2010). Effects of sample size and intraspecific variation in phylogenetic comparative studies: a meta-analytic review. *Biological Reviews, 85*, 797-805. doi:10.1111/j.1469-185X.2010.00126.x

Garland, T., Bennett, A. F., & Rezende, E. L. (2005). Phylogenetic approaches in comparative physiology. *Journal of Experimental Biology, 208*, 3015-3035. doi:10.1242/jeb.01745

Garland, T., Harvey, P. H., & Ives, A. R. (1992). Procedures for the Analysis of Comparative Data Using Phylogenetically Independent Contrasts. *Systematic Biology, 41*, 18-32. doi:10.1093/sysbio/41.1.18

Gittleman, J. L. (1986). Carnivore life history patterns: allometric, phylogenetic, and ecological associations. *American Naturalist, 127*, 744-771.

Gittleman, J. L. (1989). The comparative approach in ethology: aims and limitations. In P. P. G. Bateson & P. H. Klopfer (Eds.), *Perspectives in ethology* (Vol. 8, pp. 55-83). New York: Plenum Press.

Gittleman, J. L., & Luh, H.-K. (1992). On Comparing Comparative Methods. *Annual Review of Ecology and Systematics, 23*, 383-404.

Gonzalez-Voyer, A., & Von Hardenberg, A. (2014). An introduction to phylogenetic path analysis *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 201-229): Springer.

Goolsby, E. W., Bruggeman, J., & Ané, C. (2016). Rphylopars: fast multivariate phylogenetic comparative methods for missing data and within-species variation. *Methods in Ecology and Evolution*, *8,* 22-27. doi:10.1111/2041-210X.12612

GOV.UK. (2013). *Animal welfare legislation: protecting pets*. Retrieved from London: https://www.gov.uk/guidance/animal-welfare-legislation-protecting-pets#animal-welfare-act-2006

Graber, S. (2013). *Phylogenetic comparative methods for discrete responses in evolutionary biology.* (Masters), Zurich.

Grafen, A. (1989). The Phylogenetic Regression. *Philosophical Transactions of the Royal Society London B, 326*, 119-157.

Greco, B. J., Meehan, C. L., Hogan, J. N., Leighty, K. A., Mellen, J., Mason, G. J., & Mench, J. A. (2016). The days and nights of zoo elephants: using epidemiology to better understand stereotypic behavior of African elephants (*Loxodonta africana*) and Asian elephants (*Elephas maximus*) in North American zoos. *Plos One, 11*, e0144276.

Grimm, D. (2011). Are Dolphins Too Smart for Captivity? *Science, 332*, 526-529.

Gusset, M., Fa, J. E., & Sutherland, W. J. (2014). A horizon scan for species conservation by zoos and aquariums. *Zoo Biology, 33*, 375-380.

Hadfield, J., & Nakagawa, S. (2010). General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *Journal of Evolutionary Biology, 23*, 494-508.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software, 33*, 1-22.

Hardenberg, A. v., & Gonzalez-Voyer, A. (2013). Disentangling evolutionary cause-effect relationships with phylogenetic confirmatory path analysis *Evolution, 67*, 378-387. doi:10.1111/j.1558-5646.2012.01790.x

Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeek, M. A., Moreno-Roark, F., Near, T. J. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution, 64*, 2385-2396.

Harvey, P. H., & Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. New York: Oxford University Press.

Harvey, P. H., & Rambaut, A. (1998). Phylogenetic extinction rates and comparative methodology. *Proceedings of the Royal Society of London B: Biological Sciences, 265*, 1691-1696.

Hediger, H. (1950). *Wild Animals in Captivity*. London: Butterworths.

Hill, A. V. (1950). The Dimension of Animals and their Muscular Dynamics. *Science Progress (1933-), 38*, 209-230.

Hill, S. P., & Broom, D. M. (2009). Measuring zoo animal welfare: theory and practice. *Zoo Biology, 28*, 531-544. doi:10.1002/zoo.20276

Ho, L., & Ané, C. (2014). Package "Phylolm"-Phylogenetic Linear Regression. Retrieved from https://cran.r-project.org/web/packages/phylolm/phylolm.pdf

Hosey, G., Melfi, V., & Pankhurst, S. (2013). Working towards self-sustaining populations in captivity. In G. Hosey, V. Melfi, & S. Pankhurst (Eds.), *Zoo animals: behaviour, management, and welfare* (pp. 320). Oxford, UK: Oxford University Press.

Ives, A. R., & Garland, T. (2010). Phylogenetic logistic regression for binary dependent variables. *Systematic Biology, 59*, 9-26.

Ives, A. R., Midford, P. E., & Garland, T. (2007). Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology, 56*, 252-270.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature, 491*, 444-448. doi:http://www.nature.com/nature/journal/v491/n7424/abs/nature11631.html#supplementary-information

Jetz, W., Thomas, G. H., Joy, J. B., Redding, D. W., Hartmann, K., & Mooers, A. O. (2014). Global distribution and conservation of evolutionary distinctness in birds. *Current Biology, 24*, 919-930.

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, D. C. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. *Ecology, 90*, 2648-2648.

Kahn, C. M., & Line, S. (2007). Section 9: Special Subjects. In C. M. Kahn & S. Line (Eds.), *The Merck/Merial Manual For Pet Health: The complete health resource for your dog, cat, horse or other pets-in everyday language* (pp. 1212): Simon and Schuster.

Kaumanns, W., & Singh, M. (2015). Towards more individual-based and fitness-oriented captive mammal population management. *Current Science, 109*, 1121.

Kroshko, J. (2015). *Population-Level Risk Factors for Stereotypic Behaviour and Infant Mortality in Captive Carnivores.* (Master of Science), Guelph, Guelph, Ontario, Canada.

Kroshko, J., Clubb, R., Harper, L., Mellor, E., Moehrenschlage, A., & Mason, G. (2016). Stereotypic route-tracing in captive Carnivora is predicted by species-typical home range sizes and hunting styles. *Animal Behaviour, 117*, 197-209.

Lees, C. M., & Wilcken, J. (2009). Sustaining the Ark: the challenges faced by zoos in maintaining viable populations. *International Zoo Yearbook, 43*, 6-18. doi:10.1111/j.1748-1090.2008.00066.x

Lefebvre, L., Reader, S. M., & Sol, D. (2013). Innovating innovation rate and its relationship with brains, ecology and general intelligence. *Brain, Behavior and Evolution, 81*, 143-145.

Lefebvre, L., & Sol, D. (2008). Brains, lifestyles and cognition: are there general trends? *Brain, Behavior and Evolution, 72*, 135-144.

Lemaître, J.-F., Gaillard, J.-M., Lackey, L. B., Clauss, M., & Müller, D. W. H. (2013). Comparing free-ranging and captive populations reveals intra-specific variation in aging rates in large

herbivores. *Experimental Gerontology, 48*(2), 162-167. doi:https://doi.org/10.1016/j.exger.2012.12.004

Lemaître, J. F., Müller, D. W. H., & Clauss, M. (2014). A test of the metabolic theory of ecology with two longevity data sets reveals no common cause of scaling in biological times. *Mammal Review, 44*, 204-214. doi:doi:10.1111/mam.12023

Lewis, K. D., Shepherdson, D. J., Owens, T. M., & Keele, M. (2010). A survey of elephant husbandry and foot health in North American zoos. *Zoo Biology, 29*, 221-236. doi:doi:10.1002/zoo.20291

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing, 10*, 325-337.

Mace, G. M., & Pelletier, F. (2007). Commentary on Ricklefs & Cadena (2007): Reproductive Investment and Lifespan. *Ecology Letters, 10*, 872-874. doi:doi:10.1111/j.1461-0248.2007.01097.x

Maddison, D. R., Swofford, D. L., & Maddison, W. P. (1997). NEXUS: an extensible file format for systematic information. *Systematic Biology, 46*, 590-621.

Maddison, W. (1989). Reconstructing character evolution of polytomous cladograms. *Cladistics, 5*, 365-377. doi:10.1111/j.1096-0031.1989.tb00569.x

Maddison, W. P., & Maddison, D. R. (2011). Mesquite: a modular system for evolutionary analysis. Version 2.75. Retrieved from http://mesquiteproject.org.

Mammalogists, A. S. o. (2017). Mammal Species. Retrieved from https://academic.oup.com/mspecies

Maple, T. (1979). Great apes in captivity: The good, the bad, and the ugly. *Captivity and behavior: Primates in breeding colonies, laboratories and zoos*, 239-272.

Martins, E. P., Diniz-Filho, J. A., & Housworth, E. A. (2002). Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution, 56*, 1-13.

Martins, E. P., & Garland, T. (1991). Phylogenetic analysis of the correlated evolution of continuous characters: a simulation study. *Evolution, 45*, 534-557.

Mason, G., Burn, C. C., Dallaire, J. A., Kroshko, J., McDonald Kinkaid, H., & Jeschke, J. M. (2013). Plastic animals in cages: behavioural flexibility and responses to captivity. *Animal Behaviour, 85*, 1113-1126. doi:http://dx.doi.org/10.1016/j.anbehav.2013.02.002

Mason, G., Leipoldt, A., & de Jonge, G. (1995). *Why do female mink with high stereotypy levels have slow-growing offspring?* Paper presented at the Proceedings of the 29th International Congress of the International Society for Applied Ethology, Exeter, UK.

Mason, G., & Mendl, M. (1997). Do the stereotypies of pigs, chickens and mink reflect adaptive species differences in the control of foraging? *Applied Animal Behaviour Science, 53*, 45-58. doi:http://dx.doi.org/10.1016/S0168-1591(96)01150-1

Mason, G. J. (2010). Species differences in responses to captivity: stress, welfare and the comparative method. *Trends in Ecology & Evolution, 25*, 713-721. doi:http://dx.doi.org/10.1016/j.tree.2010.08.011

Mason, G. J. (2015). *Using Species Differences in Health and Well-being to Identify Intrinsic Risk and Protective Factors* (Vol. 16). Switzerland: WAZA.

Mason, G. J., & Veasey, J. S. (2010). What do population-level welfare indices suggest about the well-being of zoo elephants? *Zoo Biology, 29*, 256-273. doi:10.1002/zoo.20303

McDonald Kinkaid, H. Y. (2015). *Species-Level Determinants of Stereotypic Behaviour, Reproductive Success, and Lifespan in Captive Parrots (Psittaciformes).* (Doctor of Philosophy), University of Guelph, Canada.

McDougall, P. T., Réale, D., Sol, D., & Reader, S. M. (2006). Wildlife conservation and animal temperament: causes and consequences of evolutionary change for captive, reintroduced, and wild populations. *Animal Conservation, 9*, 39-48. doi:10.1111/j.1469-1795.2005.00004.x

McGowan, P. J. K., Traylor-Holzer, K., & Leus, K. (2017). IUCN Guidelines for Determining When and How Ex Situ Management Should Be Used in Species Conservation. *Conservation Letters, 10*, 361-366. doi:10.1111/conl.12285

Melfi, V. A. (2009). There are big gaps in our knowledge, and thus approach, to zoo animal welfare: a case for evidence-based zoo animal management. *Zoo Biology, 28*, 574-588. doi:doi:10.1002/zoo.20288

Mellor, D. J., Hunt, S., & Gusset, M. (2015). *Caring for Wildlife: the World Zoo and Aquarium Animal Welfare Strategy*. Retrieved from Gland, Switzerland:

Mellor, E. (2014). *Can we use biological risk factors for stereotypic behaviour in parrots to predict husbandry risk factors?* (Master of Science), Plymouth University, Plymouth, UK.

Mellor, E., Mendl, M., Bandeli, M., Cuthill, I., & Mason, G. (2017). *Abnormal repetitive route-tracing in captive Carnivora: is natural foraging niche a risk factor?* Paper presented at the Behaviour, 35th International Ethological Conference & ASAB Summer Meeting, Estoril, Portugal.

Midford, P., Garland, T., & Maddison, W. (2010). PDAP: PDTREE: A translation of the PDTREE application of Garlands *et al*.'s Phenotypic Diversity Analysis Programs. Retrieved from http://mesquiteproject.org/pdap_mesquite/

Miller, L. J. (2012). Visitor reaction to pacing behavior: influence on the perception of animal care and interest in supporting zoological institutions. *Zoo Biology, 31*, 242- 248.

Miller, M. A., Hogan, J. N., & Meehan, C. L. (2016). Housing and Demographic Risk Factors Impacting Foot and Musculoskeletal Health in African Elephants (*Loxodonta africana*) and Asian Elephants (*Elephas maximus*) in North American Zoos. *Plos One, 11*, e0155223. doi:10.1371/journal.pone.0155223

Müller, D. W., Lackey, L. B., Streich, W. J., Fickel, J., Hatt, J. M., & Clauss, M. (2011). Mating system, feeding type and ex situ conservation effort determine life expectancy in captive ruminants. *Proceedings of the Royal Society B: Biological Sciences, 278*, 2076-2080. doi:10.1098/rspb.2010.2275

Müller, D. W. H., Lackey, L. B., Streich, W. J., Hatt, J. M., & Clauss, M. (2010). Relevance of management and feeding regimes on life expectancy in captive deer. *American Journal of Veterinary Research, 71*, 275-280.

Munson, L. (1993). Diseases of captive cheetahs (*Acinonyx jubatus*): Results of the cheetah research council pathology survey, 1989–1992. *Zoo Biology, 12*, 105-124. doi:doi:10.1002/zoo.1430120110

O'Regan, H. J., & Kitchener, A. C. (2005). The effects of captivity on the morphology of captive, domesticated and feral mammals. *Mammal Review, 35*, 215-230. doi:10.1111/j.1365-2907.2005.00070.x

Orme, D. (2013). The caper package: comparative analysis of phylogenetics and evolution in R. *R package version.* Retrieved from https://cran.r project.org/web/packages/caper/vignettes/caper.pdf

Overington, S. E., Morand-Ferron, J., Boogert, N. J., & Lefebvre, L. (2009). Technical innovations drive the relationship between innovativeness and residual brain size in birds. *Animal Behaviour, 78*, 1001-1010. doi:http://dx.doi.org/10.1016/j.anbehav.2009.06.033

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature, 401*, 877-884.

Pagel, M. D. (1992). A method for the analysis of comparative data. *Journal of Theoretical Biology, 156*, 431-442. doi:http://dx.doi.org/10.1016/S0022-5193(05)80637-X

Pagel, M. D., & Meade, A. (2014). BayesTraits version 2.0 (Beta). Retrieved from http://www.evolution.rdg.ac.uk/BayesTraits.html

Paradis, E. (2012). *Analysis of Phylogenetics and Evolution with R*. New York: Springer Science & Business Media.

Paradis, E., & Claude, J. (2002). Analysis of Comparative Data Using Generalized Estimating Equations. *Journal of Theoretical Biology, 218*, 175-185. doi:citeulike-article-id:2306360

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics, 20*, 289-290.

Peng, J., Jiang, Z., Qin, G., Huang, Q., Li, Y., Jiao, Z., Zhang, F., Li,, Z., Zhang, J., Lu, Y., Liu, X., Liu, J. (2007). Impact of activity space on the reproductive behaviour of giant panda (*Ailuropoda melanoleuca*) in captivity. *Applied Animal Behaviour Science, 104*, 151-161. doi:https://doi.org/10.1016/j.applanim.2006.04.029

Plowman, A. B., Hosey, G., & Stevenson, M. (2006). *Zoo Research Guidelines: Surveys and Questionnaires*. London: BIAZA.

Pomerantz, O., Meiri, S., & Terkel, J. (2013). Socio-ecological factors correlate with levels of stereotypic behavior in zoo-housed primates. *Behavioural Processes, 98*, 85-91. doi:10.1016/j.beproc.2013.05.005

Prescott, M. J., & Buchanan-Smith, H. M. (2004). Cage sizes for tamarins in the laboratory. *Animal Welfare, 13*, 151-158.

Purvis, A., & Rambaut, A. (1995). Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Computer Applications in the Biosciences: CABIOS, 11*, 247-251. doi:10.1093/bioinformatics/11.3.247

Quadros, S., Goulart, V. D. L., Passos, L., Vecci, M. A. M., & Young, R. J. (2014). Zoo visitor effect on mammal behaviour: Does noise matter? *Applied Animal Behaviour Science, 156*, 78-84. doi:https://doi.org/10.1016/j.applanim.2014.04.002

Quick, D. L. F. (1984). An integrative approach to environmental engineering in zoos. *Zoo Biology, 3*, 65-77. doi:10.1002/zoo.1430030107

Revell, L. J. (2010). Phylogenetic signal and linear regression on species data. *Methods in Ecology and Evolution, 1*, 319-329.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution, 3*, 217-223.

Rezende, E. L., Bozinovic, F., & Garland, T. J. (2004). Climatic adaptation and the evolution of basal and maximum rates of metabolism in rodents. *Evolution, 58*, 1361-1374. doi:10.1111/j.0014-3820.2004.tb01714.x

RSPCA. (2018). RSPCA Homepage. Retrieved from https://www.rspca.org.uk/home

Schmid, J., Heistermann, M., Ganslosser, U., & Hodges, J. (2001). Introduction of foreign female Asian elephants (*Elephas maximus*) into an existing group: behavioural reactions and changes in cortisol levels. *Animal Welfare, 10*, 357-372.

Schmid, R., Doherr, M. G., & Steiger, A. (2006). The influence of the breeding method on the behaviour of adult African grey parrots (*Psittacus erithacus*). *Applied Animal Behaviour Science, 98*, 293-307. doi:http://dx.doi.org/10.1016/j.applanim.2005.09.002

Schneider, M., Nogge, G., & Kolter, L. (2014). Implementing unpredictability in feeding enrichment for Malayan sun bears (*Helarctos malayanus*). *Zoo Biology, 33*, 54-62. doi:10.1002/zoo.21112

Shepherdson, D., Lewis, K. D., Carlstead, K., Bauman, J., & Perrin, N. (2013). Individual and environmental factors associated with stereotypic behavior and fecal glucocorticoid metabolite levels in zoo housed polar bears. *Applied Animal Behaviour Science, 147*, 268-277. doi:https://doi.org/10.1016/j.applanim.2013.01.001

Siroski, P. A., Poletta, G. L., Fernandez, L., Ortega, H. H., & Merchant, M. E. (2012). Ultraviolet Radiation on Innate Immunity and Growth of Broad-Snouted Caiman (*Caiman latirostris*): Implications for Facilities Design. *Zoo Biology, 31*, 523-533. doi:10.1002/zoo.20417

Snyder, R. J., Perdue, B. M., Powell, D. M., Forthman, D. L., Bloomsmith, M. A., & Maple, T. L. (2012). Behavioral and hormonal consequences of transporting giant pandas from China to the United States. *Journal of Applied Animal Welfare Science, 15*, 1-20.

Species360. (2018). Data science for zoos and aquariums. Retrieved from https://www.species360.org/products-services/zoo-aquarium-animal-management-software/

Struck, K., Videan, E. N., Fritz, J., & Murphy, J. (2007). Attempting to reduce regurgitation and reingestion in a captive chimpanzee through increased feeding opportunities: a case study. *Lab Animal, 36*, 35.

Su, S., Cassey, P., Vall-llosera, M., & Blackburn, T. M. (2015). Going Cheap: Determinants of Bird Price in the Taiwanese Pet Market. *Plos One, 10*, e0127482. doi:10.1371/journal.pone.0127482

Symonds, M. R., & Blomberg, S. P. (2014). A primer on phylogenetic generalised least squares *Modern phylogenetic comparative methods and their application in evolutionary biology* (pp. 105-130): Springer.

Team, R. C. (2015). R: A language and environment for statistical computing. Retrieved from www.R-project.org/

TreeBASE. (2016). A Database of Phylogenetic Knowledge. Retrieved from https://treebase.org/treebase-web/home.html

van der Bijl, W. (2017). phylopath: Easy phylogenetic path analysis in R. *bioRxiv*, 212068. doi://doi.org/10.1101/212068

Veltman, C. J., Nee, S., & Crawley, M. J. (1996). Correlates of Introduction Success in Exotic New Zealand Birds. *The American Naturalist, 147*, 542-557.

Wallace, E. K., Kingston-Jones, M., Ford, M., & Semple, S. (2013). An investigation into the use of music as potential auditory enrichment for moloch gibbons (*Hylobates moloch*). *Zoo Biology, 32*, 423-426. doi:10.1002/zoo.21074

WAZA. (2018). Animal Welfare. Retrieved from http://www.waza.org/en/site/conservation/animal-welfare-1471340294

Westoby, M., Leishman, M., & Lord, J. (1995). Further remarks on phylogenetic correction. *Journal of Ecology, 83*, 727-729.

Williams, R. L., Porter, S. K., Hart, A. G., & Goodenough, A. E. (2012). The Accuracy of Behavioural Data Collected by Visitors in a Zoo Environment: Can Visitors Collect Meaningful Data? *International Journal of Zoology, 13,* 1-13. doi:10.1155/2012/724835

WoS. (2018). Web of Science record for: Felsenstein (1985) Phylogenies and the Comparative Method. Retrieved from http://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=D53QdNmNhh3KU1iPHa4&page=1&doc=7

Table 1. Table outlining pre-existing published but as yet untested hypotheses relevant to zoo animal welfare. All of these could be addressed using PCMs.

Hypothesis	Taxa	Prediction	Measurable predictor variable(s) for testing the prediction
I. Attributes related to ecological or behavioral plasticity:			
Ecological generalism preadapts species to good welfare in captivity (Mason *et al.,* 2013)	All	Generalists should have better welfare[†] in captivity than specialists	Number of habitats found in; geographical range; latitudinal range
Resilience / adaptability to environmental change preadapts species to good welfare in captivity (Mason *et al.,* 2013)	All	Species that can cope with sudden environmental change in the wild should have better welfare[†] in captivity than species which cannot	Whether or not species persist/thrive when exposed to urbanization; whether or not invasive (corrected for 'propagule effects', *sensu* Veltman, Nee & Crawley, 1996); whether or not thrive after reintroduction attempts (again, corrected for 'propagule effects', Veltman *et al*.,

			1996)
Species with little cognition-mediated behavioral plasticity will be too inflexible to adjust to captivity (Maple, 1979)	All	Less intelligent species should have poorer welfare† in captivity than more intelligent species	Relative brain volume; measures of behavioral innovation rates (from *e.g* Ducatez *et al.,* 2015; Lefebvre, Reader, & Sol, 2013; Lefebvre & Sol, 2008)
Captive environment is too unstimulating ('boring') for intelligent species (Grimm, 2011; Maple, 1979)	Primates (Maple, 1979) & Cetacea (Grimm, 2011)	More intelligent species should have poorer welfare† in captivity than less intelligent species	As above
II. Attributes related to being a prey species:			
Prey species hide their pain from observers (Kahn and Line, 2007)	All	Prey species should be at higher risk of what human carers perceive as sudden	Whether or not prey species as adults; whether typically predated by sight /auditory cues

		instances of severe illness or death; while non-prey species have longer periods of detectable clinical illness	
Fear of humans predisposes species to poor welfare in captivity (Hediger, 1950)	All	Bold species should have better welfare† in captivity than timid species	Flight distance from stressors (especially humans); whether or not species is stressed by ecotourism (Mason, 2010)
III. Attributes related to ranging behavior:			
Restrictions on traveling long distances compromise welfare (Couquiaud, 2005)	Cetacea	Coastal (shallow water), little-ranging species should have better welfare† in aquaria than open-ocean pelagic (deep water) species	Daily/annual distances traveled; maximum distances traveled from coast to open-ocean; coastal versus pelagic
Restrictions on ranging and/or	Psittaciformes	Little-ranging, resident species should	Daily distances traveled (including for species

migration compromise welfare (Mason *et al.,* 2013; Mason, 2010)	(but potentially all avians)	have better welfare in captivity† than widely-ranging and/or migratory ones	that do not fly); migratory *versus* resident.
Restrictions on ranging compromise Callitrichidae welfare (Mason & Mendl, 1997)	Marmosets and tamarins	Naturally wide-ranging callitrichid species should have poorer† welfare in laboratories and zoos than naturally little-ranging species	Daily distance traveled; typical home-range size
The home-range effect on stereotypic route-tracing relates to a lack of control and/or novelty in the captive environment (Kroshko *et al*., 2016)	Carnivora	Relatively nomadic species should show more route-tracing in captivity than species whose annual range is very similar to their daily range.	Ratio of daily ranging:annual ranging; annual number of den sites used; number of habitat types typically experienced by wild individuals (Kroshko *et al*., 2016)
IV. Attributes related to natural foraging niche			
Hunting behavior is a behavioral 'need' that gives rise to stereotypic	Carnivora	Pursuit hunters should be at higher risk of route-tracing than species with other hunting style / species that do not hunt	% day spent hunting; prey chase distance; hunting style; top speed when hunting; killing methods used; gaits used during chase; eating

route-tracing (Kroshko *et al.,* 2016)		at all	patterns used post-kill (Kroshko *et al*., 2016)
Post-feeding oral stereotypic behaviors derive from localized food searching (Mason & Mendl, 1997; Mason, 2010)	Ungulata and other taxa containing patch-feeders	Species that are typically patch-feeders should be at higher risk of abnormal oral behaviours (e.g. tongue-rolling) than species that graze or browse less selectively	Patchiness of food; ratio of time spent searching: time spent consuming food
Regurgitation and reingestion relates to not being able to 'trickle-feed' as in the wild (Struck, Videan, Fritz, & Murphy., 2007)	Primates	Species that spend more time feeding in the wild should be more likely to show regurgitation and reingestion than species which naturally spend little time feeding	Time spent foraging; number of eating bouts per day
Dietary generalism preadapts species	All	Dietary generalists should have better	Degree of dietary specialism

to good welfare in captivity (Mason, 2015)		welfare† in captivity than dietary specialists	
IV. Attributes related to other types of natural behavior:			
Welfare problems relate to restricted flying in captivity (Mellor, 2014; Schmid, Doherr, & Steiger, 2006)	Psittaciformes (but potentially all avians)	Species highly reliant on flight should have poorer welfare† in captivity than species that naturally show little or no flight	Number of hours spent flying/day in the wild; size of flight muscles in wild birds; natural reliance on flight to feed; natural reliance on flight to migrate
Restrictions on arboreality compromise Callitrichidae welfare (Prescott & Buchanan-Smith, 2004)	Marmosets and tamarins	Species that naturally use higher regions of the forest canopy and/or little use the forest floor should have poorer† welfare in captivity than species that naturally use lower parts of the canopy and the forest floor	Canopy levels used; ratio of time spent on forest floor: in trees; nest height in the wild

Restrictions on diving compromise welfare in Cetacea (Couquiaud, 2005)	Cetacea	Shallow water species should have better† welfare in aquaria than deep water species	Maximum dive distance; number of dives per day; maximum time spent at maximum depth underwater

† Welfare could be assessed via *e.g.* good reproductive outputs; long lifespans; low stereotypic behaviour etc. (see text).

Table 2. Reference table summarizing some currently used PCM statistical tests (including software packages that facilitate these tests, with website links where appropriate).

Test	***Preferred/required format of outcome variable***	***Preferred/required format of predictor variables***	**Comments**
PICs	Continuous with normally distributed	Continuous, though can	Relatively easy to use

	residuals	be adapted for categorical	
Available in Mesquite (Maddison & Maddison, 2011); www.mesquiteproject.org/) using the PDAP:PDTREE module (Midford, Garland, & Maddison, 2010) www.mesquiteproject.org/pdap_mesquite/		One predictor only	Good for poorly resolved trees, *e.g.* without branch length information, and/or with polytomies (Garland *et al.,* 1992; Pagel, 1992) Easy to make simple modifications of the PIC procedure, *e.g.* apply different branch length transformations for different traits (Garland *et al.,* 1992; Rezende, Bozinovic, & Garland, 2004), or assign branch lengths to be arbitrary or all equal if branch length information is missing (Grafen, 1989; Pagel, 1992) Robust to some violations of test assumptions (Garland *et al.,* 1992)
R (Team, 2015: www.r-			Typically does not incorporate non-Brownian Motion models of evolution (but see Freckleton, 2012), but can partially correct for this by transforming

project.org/ using the packages 'ape' (Paradis, Claude, & Strimmer, 2004), 'caper' (Orme, 2013)	branch lengths (Diaz-Uriarte & Garland, 1996; Garland *et al*., 1992) Possible to account for intra-specific variation in the Contrast program in PHYLIP (Felsenstein, 2008)
BayesTraits (Pagel & Meade, 2014): www.evolution.rdg.ac.uk/BayesTraits.html	Can therefore be too conservative, overcorrecting for this assumed strong signal (Diniz-Filho & Torres, 2002), so increasing Type II error
PHYLIP (Felsenstein, 2016): www.evolution.genetics	Creates non-intuitive graphs: contrast values are arbitrary (and can be biologically impossible negative values); contrasts can also be between an extant tip species and an estimated value for an ancestral node species, or between ancestral nodes, potentially even for historically impossible variables like ancestral enclosure size or IUCN status (correct, but non-

.washington.edu/phylip.html			intuitive)
Comparative analysis by independent contrasts (CAIC) for Macs (Purvis & Rambaut, 1995): http://www.bio.ic.ac.uk/evolve/software/caic/			Best for bivariate models; more complex models can be constructed using the contrast values from several individual models with the same outcome variable and different predictor variables (Garland *et al.,* 2005) but this is time-consuming
PGLS Available in:	Continuous, though also performs well with pseudo-continuous ordinal (Graber, 2013),	Continuous, categorical Multiple predictors	Can incorporate models of evolution aside from Brownian Motion (*e.g.* in 'ape': Paradis, 2012), such as Ornstein-Uhlenbeck and Early Burst (Harmon *et al.,* 2010; Symonds and Blomberg, 2014)

R (Team, 2015; www.r-project.org/) using the packages 'ape' (Paradis *et al.,* 2004), 'caper' (Orme, 2013), 'phytools' (Revell, 2012), and 'Rphylopars' (Goolsby, Bruggeman, & Ané, 2016)	with normally distributed residuals	Thus weights for the phylogenetic signal that is actually present (Symonds and Blomberg, 2014), so not overly conservative
		For continuous variables can estimate intra-specific variation and measurement error using 'phytools' (Revell, 2012) or 'Rphylopars' (Goolsby *et al*., 2016), both in R (and based on Ives *et al.,* 2007)
		Can easily use categorical predictor variables (Grafen, 1989)
BayesTraits (Pagel and Meade, 2014): www.evolution.rdg.ac.uk/BayesTraits.html		Intuitive graphs: species datapoints are plotted, with a PGLS regression line fitted (Symonds and Blomberg, 2014)

			Working with a poorly resolved phylogeny is possible (Symonds and Blomberg, 2014), but more difficult than PICs
			Need to use parameter estimates to obtain effect sizes and confidence intervals (Symonds and Blomberg, 2014)
Phylogenetic logistic regression	Binary (Ives & Garland, 2010)	Continuous, categorical (Ives & Garland, 2010)	Gives an estimation of the strength of phylogenetic signal (Ives & Garland, 2010)
Available in: R (Team, 2015; www.r-project.org/) using the package 'phylolm' (Ho & Ané,		Multiple predictors	Like PGLS, it weights according to signal present (performing as per standard logistic regression if there is no signal: Ives & Garland, 2010)
			Estimate of signal is only accurate if the binary outcome is relatively well balanced (*i.e.* similar number of 0s and 1s); otherwise performs poorly

2014)			(Graber, 2013) Only models Ornstein-Uhlenbeck as standard (Ives & Garland, 2010) (but see: Ané, 2014)
Phylogenetic generalized estimating equations Available in: R (Team, 2015; www.r-project.org/) using the package 'ape'	Continuous or discrete (counts or frequencies), normally or non-normally distributed (*e.g.,* gamma or Poisson) (Paradis & Claude, 2002)	Continuous, categorical (Paradis & Claude, 2002) Multiple predictors (Paradis & Claude, 2002)	Good for non-normally distributed outcome data Particularly recommended for discrete data (counts or frequencies) (Paradis & Claude, 2002) The expected variance-covariance matrix structure is not really appropriate for binary data (Ives & Garland, 2010)

(Paradis *et al.,* 2004)			Increased Type I error rates under certain circumstances (perhaps due to the degrees of freedom being an estimation of the true number) (Paradis & Claude, 2002) Does not give an estimation of phylogenetic signal (Graber, 2013)
Phylogenetic generalized linear mixed models (Bayesian inference using MCMC) Available in: R (Team, 2015; www.r-	Continuous or discrete (counts or frequencies), normal and non-normally distributed, categorical, binary, ordinal Multiple outcomes	Continuous, categorical Multiple predictors	Can separates out phylogenetic and environmental effects (Hadfield, 2010): potentially particularly interesting for zoo researchers Can incorporate measurement error (Hadfield & Nakagawa, 2010) (easier for outcome variables than predictors; Hadfield, J. pers. comm.) Useful for non-normal data (Hadfield, 2010)

project.org/) (Hadfield, 2010)

using the package

'MCMCglmm'

(Hadfield, 2010)

Bayesian inference is a very different philosophy from more common frequentist statistics; users may have to invest time familiarizing themselves with this. Such differences include choosing sensible priors (though default ones are available) (Graber, 2013)

WinBugs (Lunn, Thomas, Best, & Spiegelhalter, 2000): www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/

For continuous data, only Brownian Motion is modeled (Martins, Diniz-Filho, & Housworth, 2002)

Analyses can be time-consuming (Hadfield, 2010)

Note that the R package

is faster, and more user-friendly (Hadfield, 2010)			
Phylogenetic path analysis Available in: R (Team, 2015; www.r-project.org/) using the package 'phylopath' (van der Bijl, 2017)	Continuous, with normally distributed residuals	Continuous, categorical Multiple predictors	Incorporates PGLS (to account for relatedness among species), into a model-testing procedure to unpick the most likely directional, causal relationships between predictors (Gonzalez-Voyer & von Hardenberg, 2014) Using standardized path coefficients, can compare the relative strength of each causal relationship in a model (Gonzalez-Voyer & von Hardenberg, 2014) Can incorporate various models of evolution (Gonzalez-Voyer & von Hardenberg, 2014)

If variables show strong collinearity, this can affect parameter estimation (Freckleton, 2011) and may reduce power (Gonzalez-Voyer & von Hardenberg, 2014). At lower levels, however, this method is an effective way to unravel collinearity between predictors (Gonzalez-Voyer & von Hardenberg, 2014)

Slightly less power than non-phylogenetic path analyses, except under scenarios of strong signal; offset by decreased Type I error rates, even at weak levels of signal, compared to non-phylogenetic path analyses (Gonzalez-Voyer & von Hardenberg, 2014)

Not yet possible to account for intra-specific variation (Gonzalez-Voyer & von Hardenberg, 2014)

Table 3. Benefits of factoring species-typical captive welfare into collection planning: a potentially valuable role for PCMs

Type of benefit	How achieved	Why important
Improved animal welfare	Phasing out species prone to poor welfare; prioritizing species prone to good welfare	WAZA recommends that zoos should meet animals' behavioral and physical needs (Fa *et al.,* 2011; Mellor *et al*.,2015). Some welfare problems (*e.g.* manifest in stereotypic behavior and infant mortality) negatively affect public perception (*e.g.* Miller, 2012)
Efficient use of space, funds, and other resources	Prioritizing species already pre-disposed to good welfare, obviates needs for extensive research or resource allocation into improving welfare	Zoos have limited space and funds (*e.g*. Conde *et al.,* 2013; Fa *et al.,* 2011; Gusset *et al*., 2014; Lees & Wilcken, 2009; McGowan, Traylor-Holzer, & Leus, 2017)
Easy to maintain viable, self-sustaining populations	Species intrinsically prone to good welfare can readily be held in many zoos, and easily	Large effective population sizes are important for genetic viability (*e.g*. Conde *et al.,*

	achieve large effective populations sizes and self-sustainability	2013; Fa *et al.,* 2011; Lees & Wilcken, 2009); captive populations should also be net sources of animals, not net sinks (*e.g*. Clubb *et al.,* 2009; Fa *et al.*, 2011; McGowan *et al.,* 2017)
Reduced domestication	Species in which most individuals successfully reproduce are under weaker selection pressures than species in which only a small fraction of individuals do	If the population kept in zoos are to represent those in the wild, genetic adaptation to captivity is undesirable (Fa *et al.*, 2011; McDougall, Réale, Sol, & Reader, 2006; O'Regan & Kitchener, 2005)

Figure legends

Figure 1. Hypothetical example of an investigation between species-typical biology and species-typical welfare.

In this hypothetical example, a researcher wishes to investigate the relationship between reliance on flight (daily flight time) in the wild, and signs of poor captive welfare. **A)** At first glance, there appears to be a positive correlation between the two (though one would like a larger sample size than eight; see text). This might lead the researcher to naively conclude that being heavily reliant on flight is a risk factor for poor welfare in birds. **B)** However, here it is clear that rather than eight independent datapoints, we have two clusters of closely related species, so effectively reducing our sample size to just two. Before inferring any relationship between daily flight time and welfare, one must first control for non-independence amongst the species by statistically accounting for phylogeny (see text).

Photo credits: www.pixabay.com

Figure 2. Example of a phylogenetic tree.

This figure depicts a basic tree, a branching diagram depicting hypothesized relationships between these four species (Baum, 2008). From the base, the 'root' represents a common ancestor from which all species on the tree are descended. A 'node' is a split between branches, representing a speciation event wherein two daughter species (typically) evolve from a parent, ancestral species. Occasionally, three or more daughter species branch from a node: a 'polytomy' (an example is shown here, as the parent species of A and B, and species C and D). Polytomies can be 'soft', reflecting uncertainty about order of divergence, or 'hard' representing a genuine multiple speciation event (Maddison, 1989). The lines connecting the nodes are 'branches'. When branch 'lengths' are provided, they represent distance in evolutionary time since species spilt (Baum, 2008). The ends of the terminal

branches are the 'tips' of the tree: each one corresponds to a species, and their ordering and arrangement is the tree's 'topology' (Baum, 2008). Note that this tree is 'ultrametric' i.e. all tips are the same distance from the root). The structure of the tree, and thus relatedness of the species, is factored in during PCM analyses (as discussed in the "Data analysis and interpretation" section).

Figure 3. Worked example of PICs calculations.

In this hypothetical phylogeny **(A)**, squares and diamonds represent extant species, and circles represent their ancestors. Values for two attributes of interest, X and Y, are shown for each species (these are calculated rather than measured directly for ancestors, shown in italics). For a particular attribute, *differences* between the values of that attribute for pairs of related species represent phylogenetically independent datapoints, or *contrasts* (*e.g.*, the difference between X values for the pair of species represented by squares is one contrast; and the difference between X values for the pair represented by diamonds is another).

As shown in **(B)**, the two sets of contrasts (one for each of the two attributes) can be plotted against each other in order to determine whether the traits are correlated, *independent of phylogeny* (calculations for each contrast are shown in gray along the dotted lines). In this example, the greater the value of the X contrast, the greater the value of the Y; this suggests the traits are positively correlated. Note that more contrasts would clearly be needed for statistical testing, and that "raw" contrasts like these would need to be *standardized* by dividing them by their standard deviations (square root of the sum of the branch lengths between the species) before analysis via conventional statistics. The standardization effectively corrects for the degree of expected phenotypic divergence given the passage of time and a Brownian Motion model of evolution (see text).

Panels (A-B) are adapted from Figure 2 in Clubb & Mason, 2004.